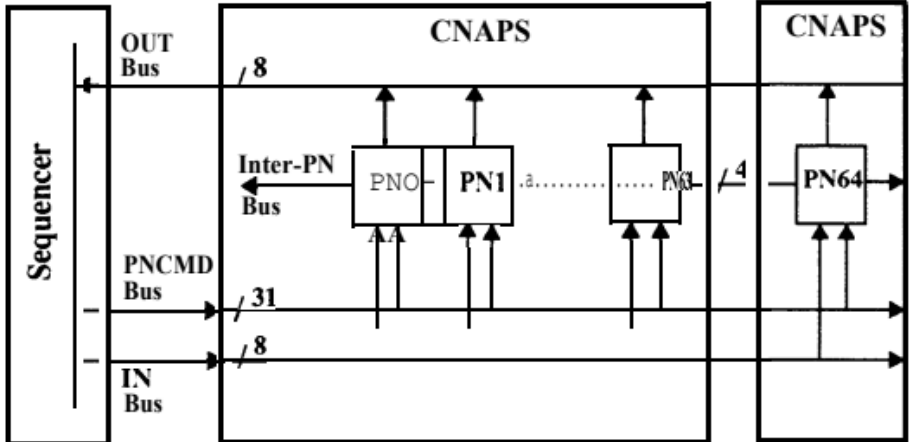# EXHIBIT F

Exhibit 3 - CNAPS
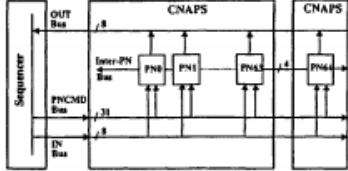
'156 Patent

Claim Limitation (Claim 7)	Exemplary Disclosure
[**156a**] A device comprising:	Connected Network of Adaptive Processors ("CNAPS") discloses a device. Specifically, the CNAPS is computer architecture consisting of an array of processors controlled by a sequencer. *See, e.g.*: "The CNAPS architecture consists of an array of processors controlled by a sequencer, both implemented as a chip set developed by Adaptive Solutions. Inc. The sequencer is a one-chip device called the CNAPS Sequencer Chip (CSC). The processor array is also a one-chip device, available with either 64 or 16 processors per chip (the CNAPS- 1064 or CNAPS- 10 16). The CSC can control up to eight 1064s or 1016s, which act like ***one large device***." Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 336 (2d ed. 1995) (emphasis added). "The digital CNAPS neurocomputer (Adaptive Solutions, Inc.) has been designed as a general purpose parallel computer for artificial neural networks and image processing tasks and was introduced by Hammerstrom in 1990. CNAPS (Connected Network of Adaptive Processors) consists of a sequencer and a linear array of maximal processor nodes and is a SIMD computer (single instruction stream – multiple data stream). A single instruction bus and three data busses connect the sequencer and the processor nodes: a broadcast input bus (8 bit), a shared output bus (8 bit), and a circular inter-PN bus (2 bit) connecting neighbouring processors (figure 3 left). The sequencer broadcasts an instruction to the processor nodes single instruction stream which synchronously execute it but process the data from their local memories multiple data stream." Peter Paschke & Ralf Möller, *Simulation of Sparse Random Networks on a CNAPS SIMD Neurocomputer*, Proc. 15th IMACS World Congress, Berlin, August 1997, at 766.

1

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	FIGURE 1 The basic CNAPS architecture. CNAPS is a single instruction, multiple data (SIMD) architecture that uses broadcast input, one-dimensional interprocessor communication, and a single shared output bus. Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 338 (2d ed. 1995). "The Adaptive Solutions CNAPS architecture is embodied in a single chip digital neurocomputer with 64 processors running at 25 megahertz. All processors receive the same instruction which they conditionally execute. Multiplication and addition are performed in parallel allowing 1.6 billion inner product steps per second per chip. Each processor has a 32-bit adder, 9-bit by 16-bit multiplier (16 by 16 in two clock cycles), shifter, logic unit, 32 16-bit registers, and 4096 bytes of local memory. Input and output are accomplished over 8-bit input and output buses common to all processors. The output bus is tied to the input bus so that output of one processor can be broadcast to all others. When multiple chips are used, they appear to the user as one chip with more processors. Special circuits support finding the maximum of values held in each processor and conserving weight space for sparsely connected networks. An accompanying sequencer chip controls instruction flow, input and output." Hal McCartor, *Back Propagation Implementation on the Adaptive Solutions CNAPS Neurocomputer Chip*, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 3, 1990, at

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	1029. Dan Hammerstrom & Gary Tahara, *CNAPS (Connected Network of Adaptive Processors)*, Presentation at Hot Chips Conference, July 8, 1991, https://www.hotchips.org/wp-content/uploads/hc_archives/hc03/3_Tue/HC3.S7/HC3.7.2.pdf.
[156b] at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value,	CNAPS discloses at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value. *See, e.g.*: "The Adaptive Solutions CNAPS architecture is embodied in a single chip digital neurocomputer with 64 processors running at 25 megahertz. All processors receive the same instruction which they conditionally execute. Multiplication and addition are performed in parallel allowing 1.6 billion inner product steps per second per chip. Each processor has a 32-bit adder, 9-bit by 16-bit multiplier (16 by 16 in two clock cycles), shifter, logic unit, 32 16-bit registers, and 4096 bytes of local memory. Input and output are accomplished over 8-bit input and output buses common to all processors." Hal McCartor, *Back Propagation Implementation on the Adaptive Solutions CNAPS Neurocomputer Chip*, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 3, 1990, at 1029. "CNAPS is a single instruction, multiple data stream (SIMD) architecture. A SIMD computer

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	has one instruction sequencing/control unit and many processor nodes (PNs). In CNAPS, the PNs are connected in a one-dimensional array (Figure 1) in which each PN can 'talk' only to its right or left neighbors. The sequencer broadcasts each instruction plus input data to all PNs, which execute the same instruction at each clock. The PNs transmit output data to the sequencer, with several arbitration modes controlling access to the output bus. As Figure 2 suggests, each PN has a local memory,4 a multiplier, an adder/subtracter, a shifter/logic unit, a register tile,' and a memory addressing unit. The entire PN uses fixed-point, two's complement arithmetic, and the precision is 16 bits, with some exceptions. The PN memory can handle 8- or 16-bit reads or writes. The multiplier produces a 24-bit output; an 8 X 16 or 8 X 8 multiply takes one clock, and a 16 X 16 multiply takes two clocks. The adder can switch between 16- or 32-bit modes. The input and output buses are 8 bits wide, and a 16-bit word can be assembled (or disassembled) from two bytes in two clocks." Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 337 (2d ed. 1995).

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	FIGURE 2 The internal structure of a CNAPS processor node (PN). Each PN has its own storage and arithmetic capabilities. Storage consists of 4096 bytes. Arithmetic operations include multiply, accumulate, logic, and shift. All units are interconnected by two 16-bit buses. Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 338 (2d ed. 1995). "The X1 system consists of a linear array of PNs. Each PN is a simple arithmetic processor with its own local memory. The array is sequenced by a single controller, thus each PN executes the same instruction at each clock, using SIMD processing. Each PN is connected to three global buses: InBus—the data input bus, PnCmd—the command bus, which indicates what operations a PN performs each clock, and OutBus - the output bus." Dan Hammerstrom, *A*

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	VLSI architecture for high-performance, low-cost, on-chip learning, 1990 IJCNN International Joint Conference on Neural Networks, 1990, vol.2, at 539. "Each Processor Node has internal units connected via control signals and several data buses. See Figure 5. The InBus and PnCmd bus enter the PN through the Input Unit. In general the PN is horizontally microcoded: 1. *Input Unit*: The Input Unit decodes the PnCmd bus and routes it to the other units, and receives an 8 bit value from the InBus. It also contains a flag to allow conditional instruction execution of each PN. Although the input and output buses are 8 bits, the internal buses are 16 bits. The Input Unit can assemble two 8 bit quantities into a 16 bit value when needed. 2. *Logic-shifter*: The Logic-shifter contains both a shifter and a logic operation unit. Both operate on 16 bit quantities. These allow computed quantities to be manipulated by shifting and bit masking. There is also a 1's counter that sums the bits in the byte and whose output is used with 1 bit weights (8 per byte) to accumulate the AND of an 8 bit input and an 8 bit weight. This permits binary inputs and binary weights (8 per byte) - 8 are computed each clock. 3. *Registerfile*: The register file contains 32 16-bit registers for intermediate storage of constants such as learning rates and the PN ID. 4. *Multiplier*: The Multiplier unit contains a 9x16 bit, two's complement multiplier. The mul- tiplier produces a 24 bit output. Using mode bits, the personality of the multiplier can be modified to multiply a positive 8 bit number by a signed 16 bit number, a signed 8 bit number by a signed 16 bit number, or a signed 16 bit number by a signed 16 bit number (this operation takes two clocks). There is a direct path from the multiplier to the adder that allows for the PN to be operated in Vector Mode where a simultaneous multiply and accumulate can occur each clock 5. *Adder*: The adder/subtractor takes 32 bit inputs and produces a two's complement 32 bit result. Adder overflow causes saturation to the largest positive or negative number, depending on the sign of the final result. 6. *Weight Address Generation*: The separate multiplier-to-adder bus allows the PN to be op- erated in a vector mode, which requires a regular stream of addresses to be generated for the weight memory. This unit contains its own adder for adding the contents of the stride

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	register to the current weight index. As a result there is arbitrary striding through memory for those programs that have more complex data structures. 7. *Weight Memory*: Memory can be accessed in either 8 bit or 16 bit mode. There is also a hardware system, called the Virtual Zeros mechanism4, which is used for the efficient representation of sparse connectivity. This is a simple set of bit mapped registers that allow groups of contiguous zeros in memory to be removed from real memory. 8. *Output Buffer*: The Output Buffer contains the arbitration logic for access to the Inter-PN bus and the OutBus. Both 16 and 8 bit values can be transmitted. 16 bit values require 2 clocks to transmit over the 8 bit OutBus. The OutBus arbitration and transmission mechanism operates separately from, but is synchronized with, the SIMD control to allow PNs to both transmit and do multiply-accumulation simultaneously. This capability is required when outputs of a layer are fed back as inputs to the next layer. This feature is called the Virtual PN5. There are several OutBus arbitration modes." Dan Hammerstrom, *A VLSI architecture for high-performance, low-cost, on-chip learning*, 1990 IJCNN International Joint Conference on Neural Networks, 1990, vol.2, at 539. "Each PN is similar to an individual Digital Signal Processor (DSP). Figure 2 shows a block diagram of a PN in the CNAPS architecture. There are eight functional units in a PN: 1. The Input unit receives data from the INbus. 2. The Output unit sends data to the OUThus. 3. The Multiplier can perform an 8x8 multiply, ai 8x16 multiply, or a 16x16 multiply. 4. The Adder can perfonn a 32x32 addition. 5. The Logic unit can perform the Boolean logic functions ofAND, OR, XOR. The logic unit can also shift words right and left. 6. The Register unit contains 32 - 16 bit registers for temporary value storage. 7. There is memory local to each PN. 8. The Memory Address Unit accesses the local memory." Thomas E. Baker, *Artificial neural network and image processing using the Adaptive Solutions' architecture*, Proc. SPIE 1452, Image Processing Algorithms and Techniques II, June 1, 1991, at 503. "The CNAPS-1064 component is a high-performance processor consisting of 64 individual

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	processor nodes (PNs) that operate in parallel. Each PN is a fully functional digital processor with an adder, multiplier, logic unit, local memory, a set of input and output buses, two internal buses that connect functional units, and an inter-PN bus that connects the PNs serially." Adaptive Solutions, Inc., *CNAPS Data Book*, CNAPS Hardware Series, October 15, 1993, at HAMMERSTROM-00000158. "The CNAPS PN array is the processing engine for a CNAPS system. The array consists of one or more CNAPS chips, each containing up to 64 processors (PNs). Each PN is a complete fixed-point processor with its own 4KB of on-chip memory. A PN can perform 1-, 8-, or 16-bit integer arithmetic and can execute a multiply-and-accumulate operation in one clock cycle. The effect is similar to having 64 digital signal processors (DSPs) on one chip." Adaptive Solutions, Inc., *Getting Acquainted with CNAPS*, CNAPS Core Series, October 15, 1993, at HAMMERSTROM-00000490.
[**156c**] wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical	As explained below and in the Responsive Contentions Regarding Non-Infringement and Invalidity ("Responsive Contentions"), it would have been obvious to one of skill in the art based on the disclosures in CNAPS (alone or in combination with the reduced precision floating point teachings of Dockser, Tong, Belanovic / Belanovic and Leeser, Lee, Shirazi, Aty, Sudha, and TMS320C32, or the logarithmic format disclosed in GRAPE-3 and Hoefflinger) that the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input. CNAPS was designed to perform neural network algorithms at the minimum required precision. "Another major simplification of the CNAPS architecture, which is found in other digital ANN chips, was the use of limited-precision, fixed-point arithmetic. Many researchers have shown that floating point and high precision are unnecessary in ANN simulation (Fahlman and

8

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
values of that same input; and	Hoehfeld, 1992). CNAPS supported 1-, 8-, and 16-bit precision in hardware. Consequently, the PNs were smaller and cheaper. This reduced precision was more than adequate for the applications implemented on CNAPS." Dan Hammerstrom, *Digital VLSI for Neural Networks*, at 9-10, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.7541&rep=rep1&type=pdf. Reduced precision computations were commonly used in connection with neurocomputers like the CNAPS. *See, e.g.*: "BP implementations typically use 32-bit floating point math. This largely eliminates scaling, precision and dynamic range issues. Efficient hardware implementation dictates integer arithmetic units with precision no greater than required. Baker [Bak90] has shown 16-bit integer weights are sufficient for BP training and much lower values adequate for use after training." Hal McCartor, *Back Propagation Implementation on the Adaptive Solutions CNAPS Neurocomputer Chip*, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 3, 1990, at 1029. "The inherent parallelism in ANN and connectionist models suggests an opportunity to speed up the simulations. Their simple, low precision computations also suggest an opportunity to employ simpler and cheaper, low-precision digital hardware implemented by full-custom silicon or by FPGAs (Field Programmable Gate Arrays)." Dan Hammerstrom, *Digital VLSI for Neural Networks*, at 2, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.7541&rep=rep1&type=pdf. "Conforms to idea that [Artificial Neural Network] consists of large numbers of simple, low precision processing elements. . . NetTalk, for example, showed no appreciable difference between 32 bit floating point and 15 bit integer." Dan Hammerstrom & Gary Tahara, *CNAPS (Connected Network of Adaptive ProcessorS)*, Presentation at Hot Chips Conference, July 8, 1991, at 3, https://www.hotchips.org/wp-content/uploads/hc_archives/hc03/3_Tue/HC3.S7/HC3.7.2.pdf. "The limited precision simulation results show that sixteen total bits of precision (one sign bit,

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	three bits to the left of the binary point and twelve bits to the right) are adequate. . . . A floating point representation may be able to learn successfully with fewer than twelve bits in the mantissa, but floating point units are too expensive for massively parallel architectures." Tom Baker & Dan Hammerstrom, *Modifications to Artificial Neural Networks Models for Digital Hardware Implementation*, Oregon Graduate Center, Technical Report No. CS/E 88-035, December 1988, https://scholararchive.ohsu.edu/downloads/pc289j41v?locale=en. Accordingly, CNAPS discloses the use of arithmetic units designed to perform calculations using reduced-precision 1-, 8-, and 16-bit math for calculations that typically use 32-bit floating point math, operating on only the 1-, 8-, or 16 most significant bits in the registers. *See, e.g.*: "The X1 has three basic weight modes: 1 bit, 8 bits (7 mantissa + sign), and 16 bits (15 mantissa + sign). In 8 and 16 bit modes, there is a sign bit. When the mantissa is zero, it is generally considered a null weight. Most algorithms require 8 bits, but there are some algorithms that need only single bit precision and some that require up to 16 bits. Bit mode does not have an explicit sign but can be run twice with both positive and negative weight vectors. Two's complement representation is used throughout. Many of the units recognize and deal with negative numbers appropriately, other units assume unsigned values, that is, the sign bit is ignored. Values transmitted onto and off the chip can be either 8 or 16 bits (16 bit values require two uses of the 8 bit buses). The preferred mode is 8 bits, since for most neural network algorithms that is sufficient. The 16 bit mode is provided for those applications, such as digital signal processing or the loading of 16 bit weights, that need more precision. With 16 bit values, bit 15 is always the sign bit. 8 bit values can be switched to signed or unsigned." Dan Hammerstrom & Hal McCartor, *X1 Programmer's Guide and Reference Manual*, Adaptive Solutions, Inc., March 7, 1990, at HAMMERSTROM-00000276. "Each PN provides four data precisions for representing data: 1-bit, 8-bit (signed or unsigned), 16-bit (signed or unsigned) and 32-bit (signed). Most algorithms use 8 bits, but some algorithms need only single-bit precision and some require up to 16 bits. Single-bit data does not have an explicit sign, but the network can be run twice with both positive and negative data

10

Exhibit 3 - CNAPS

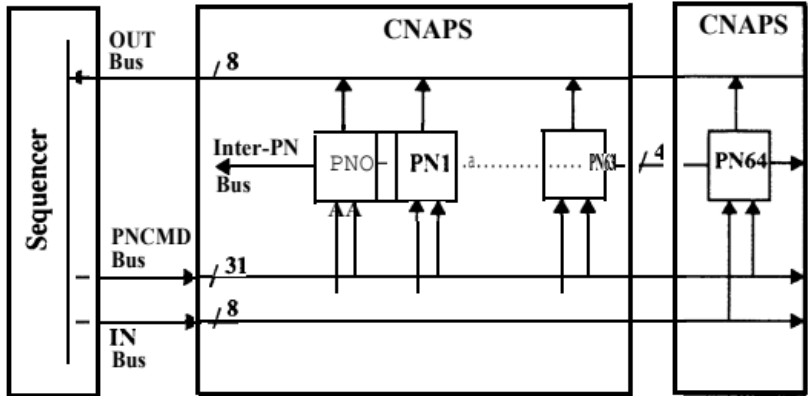| Claim Limitation (Claim 7)	Exemplary Disclosure
	values." Adaptive Solutions, Inc., *CNAPS Data Book*, CNAPS Hardware Series, October 15, 1993, at HAMMERSTROM-00000154. CNAPS further discloses expanding the dynamic range of the possible valid inputs to an operation by reducing the precision of the operation's calculations. "In place of floating point variables, the CNAPS provides scaled variables, of one or two bytes, in which a fixed point can be set at a given bit to separate the whole number part from the fraction. Choosing the fixed point position is an important consideration. A low fixed point reduces the precision, whereas a high fixed point to achieve greater precision may result in overflows of the whole part. CNAPS programs often have multiple scaled variables with different fixed points to be used where greater or less fractional precision is necessary." Jason M. Kinser and Thomas Lindblad, *Implementation of pulse-coupled neural networks in a CNAPS environment*, IEEE TRANSACTIONS ON NEURAL NETWORKS, vol. 10, no. 3, May 1999. "In the CNAPS real numbers are represented by a scaled type. The scaled value requires the user to define the number of bits that will represent the integer part of the real number and the number of bits that will represent the decimal portion of the number with the total number of bits being 16 or less. Since this is not the mantissa/exponent representation the dynamic range of the scaled type is quite limited." Jason M. Kinser and Thomas Lindblad, *Implementation of pulse-coupled neural networks in a CNAPS environment*, IEEE TRANSACTIONS ON NEURAL NETWORKS, vol. 10, no. 3, May 1999. "The CNAPS architecture uses fixed-point arithmetic, so floating-point values must be converted (or *scaled*) to fixed-point values. CNAPS-C supports fixed-point dated through the `scaled` type construction, which lets you define a fixed-point type with a specified precision." Adaptive Solutions, Inc., *Getting Acquainted with CNAPS*, CNAPS Core Series, October 15, 1993, at HAMMERSTROM-00000497. For the reasons explained in the Responsive Contentions, it would have been obvious to one of skill in the art to have substituted the fixed-point number format used in CNAPS for a floating-point format that met the claimed minimum range and precision requirements, and to have used

Exhibit 3 - CNAPS

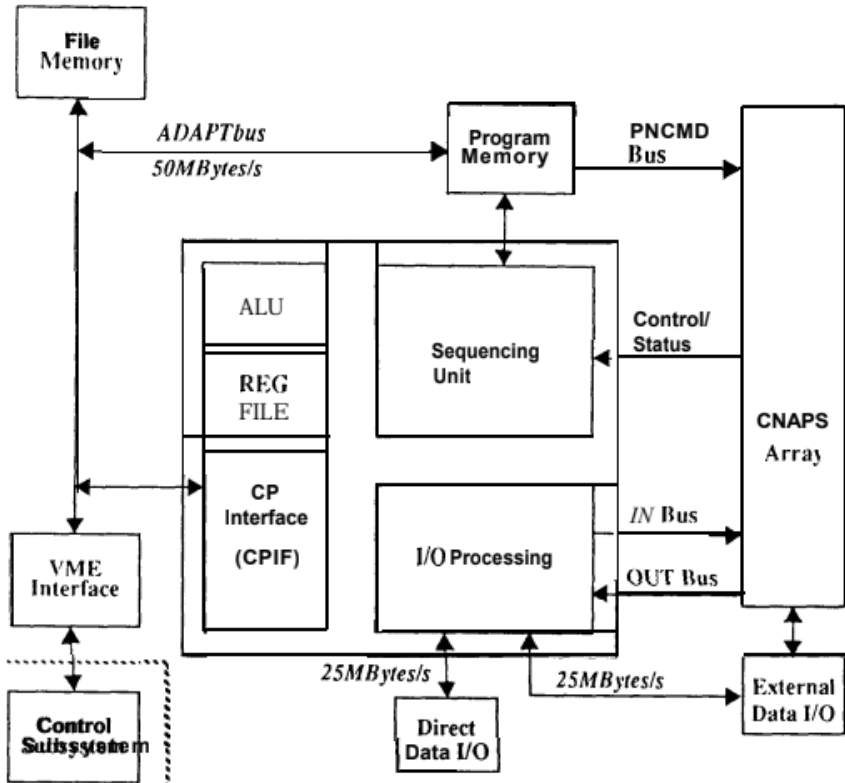| Claim Limitation (Claim 7)	Exemplary Disclosure
	the reduced-precision floating-point number formats disclosed in Tong, Dockser, Belanovic / Belanovic and Leeser, Shirazi, Sudha, Aty, and TMS 320C32 or the logarithmic format disclosed in GRAPE-3 and Hoefflinger, either alone or in combination. *See also* Appendix to Responsive Contentions (detailing error rates associated with different mantissa sizes).
[**156d**] at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit;	CNAPS discloses at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit. *See, e.g.*: "The digital CNAPS neurocomputer (Adaptive Solutions, Inc.) has been designed as a general purpose parallel computer for artificial neural networks and image processing tasks and was introduced by Hammerstrom in 1990. CNAPS (Connected Network of Adaptive Processors) consists of a sequencer and a linear array of maximal processor nodes and is a SIMD computer (single instruction stream – multiple data stream). A single instruction bus and three data busses connect the sequencer and the processor nodes: a broadcast input bus (8 bit), a shared output bus (8 bit), and a circular inter-PN bus (2 bit) connecting neighbouring processors (figure 3 left). The sequencer broadcasts an instruction to the processor nodes single instruction stream which synchronously execute it but process the data from their local memories multiple data stream." Peter Paschke & Ralf Möller, *Simulation of Sparse Random Networks on a CNAPS SIMD Neurocomputer*, Proc. 15th IMACS World Congress, Berlin, August 1997, at 766. "The Adaptive Solutions CNAPS architecture is embodied in a single chip digital neurocomputer with 64 processors running at 25 megahertz. All processors receive the same instruction which they conditionally execute. Multiplication and addition are performed in parallel allowing 1.6 billion inner product steps per second per chip. Each processor has a 32-bit adder, 9-bit by 16-bit multiplier (16 by 16 in two clock cycles), shifter, logic unit, 32 16-bit registers, and 4096 bytes of local memory. Input and output are accomplished over 8-bit input and output buses common to all processors. The output bus is tied to the input bus so that output of one processor can be broadcast to all others. When multiple chips are used, they appear to the user as one chip with more processors. Special circuits support finding the maximum of values held in each processor and conserving weight space for sparsely connected networks. An accompanying sequencer chip controls instruction flow, input and output." Hal McCartor, *Back Propagation Implementation on the Adaptive Solutions CNAPS*

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	Neurocomputer Chip, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 3, 1990, at 1029. "The CNAPS architecture consists of an array of processors controlled by a sequencer, both implemented as a chip set developed by Adaptive Solutions. Inc. The sequencer is a one-chip device called the CNAPS Sequencer Chip (CSC). The processor array is also a one-chip device, available with either 64 or 16 processors per chip (the CNAPS- 1064 or CNAPS- 10 16)." Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 336 (2d ed. 1995). "CNAPS is a single instruction, multiple data stream (SIMD) architecture. A SIMD computer has one instruction sequencing/control unit and many processor nodes (PNs). In CNAPS, the PNs are connected in a one-dimensional array (Figure 1) in which each PN can "talk" only to its right or left neighbors. The sequencer broadcasts each instruction plus input data to all PNs, which execute the same instruction at each clock. The PNs transmit output data to the sequencer, with several arbitration modes controlling access to the output bus." Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 337 (2d ed. 1995). "The CSC sequencer (Figure 3) performs program sequencing for the PN array and has private access to a program memory. The CSC also performs input/ output (l/O) processing for the array, writing input data to the array and reading output data from it." Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 337 (2d ed. 1995).

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	FIGURE 1 The basic CNAPS architecture. CNAPS is a single instruction, multiple data (SIMD) architecture that uses broadcast input, one-dimensional interprocessor communication, and a single shared output bus. Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 338 (2d ed. 1995).

14

Exhibit 3 - CNAPS

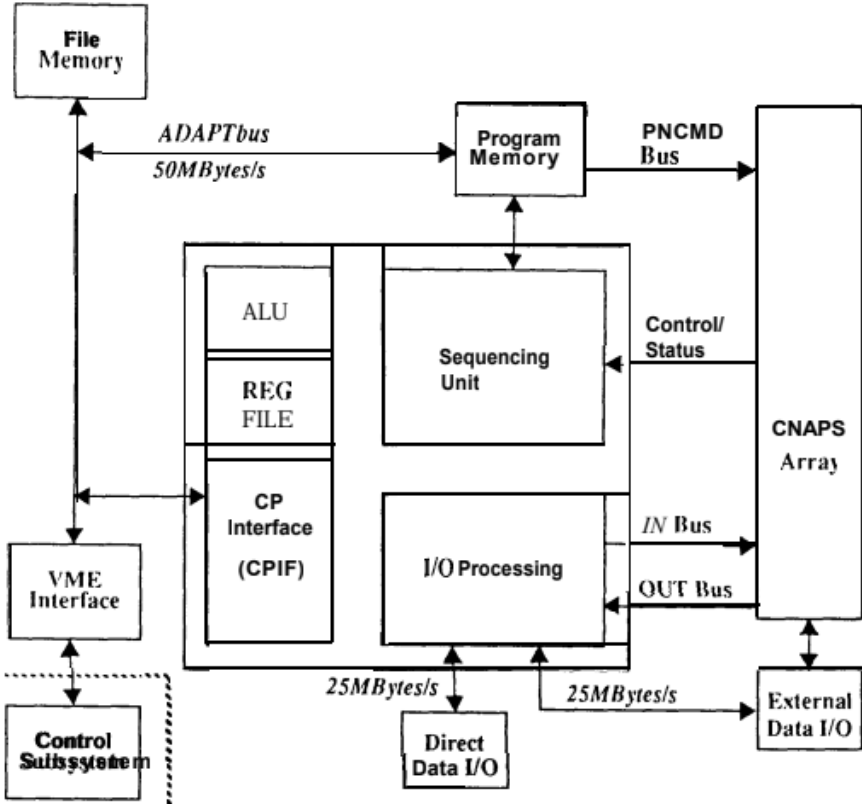| Claim Limitation (Claim 7)	Exemplary Disclosure
	FIGURE 3 The CNAPS sequencer chip (CSC) internal structure. The CSC accesses an external program store, which contains both CSC and CNAPS PN array instructions. PN array instructions are broadcast to all PNs. CSC instructions control sequencing and all array input and output. Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 339 (2d ed. 1995). "A separate sequencer containing an instruction store and miscrosequencer controls SIMD PN execution. The sequencer places data or literals onto the IN bus and forward sequencing commands to the PN array." Adaptive Solutions, Inc., *CNAPS Neurocomputing*, 1991, at HAMMERSTROM-00000999.

15

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
[**156e**] wherein the at least one first computing device comprises at least one of a central processing unit (CPU), a graphics processing unit (GPU), a field programmable gate array (FPGA), a microcode-based processor, a hardware sequencer, and a state machine;	CNAPS discloses at least one first computing device comprises at least one of a central processing unit (CPU), a graphics processing unit (GPU), a field programmable gate array (FPGA), a microcode-based processor, a hardware sequencer, and a state machine. *See, e.g.*: "The digital CNAPS neurocomputer (Adaptive Solutions, Inc.) has been designed as a general purpose parallel computer for artificial neural networks and image processing tasks and was introduced by Hammerstrom in 1990. CNAPS (Connected Network of Adaptive Processors) consists of a sequencer and a linear array of maximal processor nodes and is a SIMD computer (single instruction stream – multiple data stream). A single instruction bus and three data busses connect the sequencer and the processor nodes: a broadcast input bus (8 bit), a shared output bus (8 bit), and a circular inter-PN bus (2 bit) connecting neighbouring processors (figure 3 left). The sequencer broadcasts an instruction to the processor nodes single instruction stream which synchronously execute it but process the data from their local memories multiple data stream." Peter Paschke & Ralf Möller, *Simulation of Sparse Random Networks on a CNAPS SIMD Neurocomputer*, Proc. 15th IMACS World Congress, Berlin, August 1997, at 766. "The Adaptive Solutions CNAPS architecture is embodied in a single chip digital neurocomputer with 64 processors running at 25 megahertz. All processors receive the same instruction which they conditionally execute. Multiplication and addition are performed in parallel allowing 1.6 billion inner product steps per second per chip. Each processor has a 32-bit adder, 9-bit by 16-bit multiplier (16 by 16 in two clock cycles), shifter, logic unit, 32 16-bit registers, and 4096 bytes of local memory. Input and output are accomplished over 8-bit input and output buses common to all processors. The output bus is tied to the input bus so that output of one processor can be broadcast to all others. When multiple chips are used, they appear to the user as one chip with more processors. Special circuits support finding the maximum of values held in each processor and conserving weight space for sparsely connected networks. An accompanying sequencer chip controls instruction flow, input and output." Hal McCartor, *Back Propagation Implementation on the Adaptive Solutions CNAPS Neurocomputer Chip*, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 3, 1990, at 1029. "The CNAPS architecture consists of an array of processors controlled by a sequencer, both

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	implemented as a chip set developed by Adaptive Solutions. Inc. The sequencer is a one-chip device called the CNAPS Sequencer Chip (CSC). The processor array is also a one-chip device, available with either 64 or 16 processors per chip (the CNAPS- 1064 or CNAPS- 10 16).” Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 336 (2d ed. 1995). “CNAPS is a single instruction, multiple data stream (SIMD) architecture. A SIMD computer has one instruction sequencing/control unit and many processor nodes (PNs). In CNAPS, the PNs are connected in a one-dimensional array (Figure 1) in which each PN can “talk” only to its right or left neighbors. The sequencer broadcasts each instruction plus input data to all PNs, which execute the same instruction at each clock. The PNs transmit output data to the sequencer, with several arbitration modes controlling access to the output bus.” Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 337 (2d ed. 1995). FIGURE 1 The basic CNAPS architecture. CNAPS is a single instruction, multiple data (SIMD) architecture that uses broadcast input, one-dimensional interprocessor communication, and a single shared output bus. Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 338 (2d ed. 1995).

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	FIGURE 3 The CNAPS sequencer chip (CSC) internal structure. The CSC accesses an external program store, which contains both CSC and CNAPS PN array instructions. PN array instructions are broadcast to all PNs. CSC instructions control sequencing and all array input and output. Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 339 (2d ed. 1995).
[156f] and, wherein the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the	CNAPS discloses the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication. *See, e.g.*: "To increase the performance of the CNAPS array, multiple chips can be linked together using

18

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide.	a common sequencer. Signals between PNs that cross chip boundaries are transparent to the programmer. The programmer views a multiple chip system as if it were a monolithic array of PNs (see Figure 1). A system with eight CNAPS chips has 512 PNs, and can perform 10 Billion MAC operations per second when operating at 20 MHz." Thomas E. Baker, *Artificial neural network and image processing using the Adaptive Solutions' architecture*, Proc. SPIE 1452, Image Processing Algorithms and Techniques II, June 1, 1991, at 505. "The CNAPS architecture had 64 [processor nodes] per chip. . . . Multiple chips could be combined to create larger, more powerful systems." Dan Hammerstrom, *Digital VLSI for Neural Networks*, at 10, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.7541&rep=rep1&type=pdf. "The Adaptive Solutions CNAPS architecture chip is a general purpose neurocomputer chip. It has 64 processors, each with 4 K bytes of local memory, running at 25 megahertz." Hal McCartor, *Back Propagation Implementation on the Adaptive Solutions CNAPS Neurocomputer Chip*, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 3, 1990, at 1029. "When multiple chips are used, they appear to the user as one chip with more processors." Hal McCartor, *Back Propagation Implementation on the Adaptive Solutions CNAPS Neurocomputer Chip*, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 3, 1990, at 1029. "Multiple chips can be arbitrarily combined to create large systems." Dan Hammerstrom & Gary Tahara, *CNAPS (Connected Network of Adaptive ProcessorS)*, Presentation at Hot Chips Conference, July 8, 1991, https://www.hotchips.org/wp-content/uploads/hc_archives/hc03/3_Tue/HC3.S7/HC3.7.2.pdf. "The processor array is also a one-chip device, available with either 64 or 16 processors per chip (the CNAPS- 1064 or CNAPS- 10 16). The CSC can control up to eight 1064s or 1016s, which act like one large device." Dan Hammerstrom, AN INTRODUCTION TO NEURAL AND ELECTRONIC NETWORKS 336 (2d ed. 1995).

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	"Chip boundaries across the array are arbitrary, since the user only sees a single array of [processor nodes]. Thus [CNAPS] systems are easily scalable." Dan Hammerstrom, *A VLSI architecture for high-performance, low-cost, on-chip learning*, 1990 IJCNN International Joint Conference on Neural Networks, 1990, vol.2, at 539. "The CNAPS-1064 chip is a CMOS device with the parallel processing power of a DSP and the flexibility to program it for a wide variety of applications. Each CNAPS-1064 component contains 64 interconnected processors, called processor nodes or *PNs,* which can be programmed to operate ln parallel on a given set of data, Each PN is a simple arithmetic processor with its own local memory, adder, multiplier, logic unit, register file, and I/O buffers. Multiple CNAPS-1064 chips can be strung together to dramatically improve the speed and processing power of the system, Such a connection of CNAPS-1064 chips is called the *CNAPS array* (or *PN array*). PNs in the array are numbered sequentially starting with zero (0) and continue through to the last PN on the last CNAPS chip, The PN array is controlled by a sequencing and I/0 processing unit the CNAPS CSC (See Part l for details on the CSC.)" Adaptive Solutions, Inc., *CNAPS Data Book*, CNAPS Hardware Series, October 15, 1993, at HAMMERSTROM-00000153. "The CNAPS system is a complete parallel computer system based on a custom digital chip containing up to 64 processors. Current CNAPS systems have one to eight CNAPS chips for a total of 64 to 512 processors. Each processor can execute a multiple-and-accumulate (dot product) operation in one clock cycle, so an array of 512 processors yields up to 10.24 *billion* multiply-accumulates per second at 20 MHz. The result is similar to having up to 512 digital signal processors (DSPs) in one system. CNAPS provides the computational power needed for real-world, real-time pattern recognition problems—all in a compact, economical package." Adaptive Solutions, Inc., *Getting Acquainted with CNAPS*, CNAPS Core Series, October 15, 1993, at HAMMERSTROM-00000482. "The CNAPS hardware consists of a linear array of single-instruction, multiple-data processors that operate in parallel on a data set. Each processor operates on its own slice of the data. The

Exhibit 3 - CNAPS

Claim Limitation (Claim 7)	Exemplary Disclosure
	CNAPS-C language supports this parallelism. The number of processors is conceptually unlimited: in practice, the number of CNAPS chips determines the number of processors." Adaptive Solutions, Inc., *CNAPS-C Programming Guide,* CNAPS Series, March 15, 1996, at HAMMERSTROM-00000567-568. "The CNAPS Server has 256 processors (PNs; processing nodes) that are used to simulate neural network nodes." Adaptive Solutions, Inc., *CNAPS Neurocomputing*, 1991 at HAMMERSTROM-00001003.

Exhibit 3 - CNAPS

'273 Patent

Claim Limitation (Claim 53)	Exemplary Disclosure
[273a] A device:	CNAPS discloses a device. Specifically, the CNAPS is computer architecture consisting of an array of processors controlled by a sequencer. *See* [**156a**].
[273b] comprising at least one first low precision high-dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value,	CNAPS discloses at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first input signal representing a first numerical value to produce a first output signal representing a second numerical value. *See* [**156b**].
[273c] wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X % of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input;	For the reasons explained above and in the Responsive Contentions, it would have been obvious to one of skill in the art based on the disclosures in CNAPS (alone or in combination with the reduced precision floating point teachings of Dockser, Tong, Belanovic / Belanovic and Leeser, Lee, Shirazi, Aty, Sudha, and TMS320C32 or the logarithmic format disclosed in GRAPE-3 and Hoefflinger) that the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input. *See* [**156c**].
[273d] wherein the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide.	CNAPS discloses the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide. *See* [**156f**].

Exhibit 3 - CNAPS

'961 Patent

Claim Limitation (Claim 4)	Exemplary Disclosure
[**961a**] A device comprising:	CNAPS discloses a device. Specifically, the CNAPS is computer architecture consisting of an array of processors controlled by a sequencer. *See* [**156a**].
[**961b**] at least one first low precision high-dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value,	CNAPS discloses at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first input signal representing a first numerical value to produce a first output signal representing a second numerical value. *See* [**156b**].
[**961c**] wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=10% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.2% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input; and	For the reasons explained above and in the Responsive Contentions, it would have been obvious to one of skill in the art based on the disclosures in CNAPS (alone or in combination with the reduced precision floating point teachings of Dockser, Tong, Belanovic / Belanovic and Leeser, Lee, Shirazi, Aty, Sudha, and TMS320C32 or the logarithmic format disclosed in GRAPE-3 and Hoefflinger) that the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=10% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.2% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input. *See* [**156c**].
[**961d**] at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit.	CNAPS discloses at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit. *See* [**156d**].

23

Exhibit 3 - CNAPS

Claim Limitation (Claim 13)	Exemplary Disclosure
[**961e**] A device comprising:	CNAPS discloses a device. Specifically, the CNAPS is computer architecture consisting of an array of processors controlled by a sequencer. *See* [**156a**].
[**961f**] a plurality of components comprising:	CNAPS discloses at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first input signal representing a first numerical value to produce a first output signal representing a second numerical value. *See* [**156b**]. CNAPS discloses at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit. *See above* [**156d**].
[**961g**] at least one first low precision high-dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value,	CNAPS discloses at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first input signal representing a first numerical value to produce a first output signal representing a second numerical value. *See* [**156b**].
[**961h**] wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=10% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.2% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input.	For the reasons explained above and in the Responsive Contentions, it would have been obvious to one of skill in the art based on the disclosures in CNAPS (alone or in combination with the reduced precision floating point teachings of Dockser, Tong, Belanovic / Belanovic and Leeser, Lee, Shirazi, Aty, Sudha, and TMS320C32 or the logarithmic format disclosed in GRAPE-3 and Hoefflinger) that the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000 and for at least X=10% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X% of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.2% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input. *See* [**156c**].

24